DOCUMENT RESUME

ED 406 148 SE 059 800

AUTHOR Lomask, Michal; And Others
TITLE The Validity of Portfolio-Based Assessment of Science Teachers.
PUB DATE Mar 97
NOTE 29p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (70th, Chicago, IL, March 21, 1997).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Beginning Teachers; Elementary School Teachers; Evaluation; Intermediate Grades; Middle Schools; *Portfolio Assessment; *Science Teachers; Secondary Education; Secondary School Teachers; Teacher Education
IDENTIFIERS Connecticut

ABSTRACT

This paper describes a new portfolio-based assessment program for beginning science teachers who seek a license to teach in Connecticut's middle and high schools. This is part of the comprehensive Beginning Educator Support and Training (BEST) induction program for beginning teachers in Connecticut. The science portfolio assessment requires all beginning science teachers to plan, teach, and assess students' learning in a two-week inquiry-based unit. The completed portfolios contain four types of artifacts: (1) daily logs; (2) four videotaped segments that capture the teaching and learning in various learning settings in the same unit; (3) students' work done through the entire unit; and (4) the teacher's reflections on the quality of students' learning and his or her own teaching. Evidence gathered through various explorative studies support the job-relatedness of the assessment content and the validity of inferences made from the assessment scores. Research results also show that the portfolio-based assessment program is accessible to beginning science teachers in different school systems and has no adverse impact on any sub-group of candidates. Follow-up interviews with graduates of the program showed that the BEST program has a lasting impact on many science teachers. Contains 26 references. (Author/JRH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 406 148

The Validity of Portfolio-Based Assessment of Science Teachers

Michal Lomask
Michael Seroussi
Fie Budzinski

Connecticut State Department of Education
Bureau of Research and Teacher Assessment
P. O. Box 2219, Hartford, CT 06145

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M. Lomask

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the 1997 NARST Annual Meeting, Chicago, IL, March 1997

BEST COPY AVAILABLE

The Validity of Portfolio-Based Assessment of Science Teachers

Michal Lomask, Michael Seroussi & Fie Budzinski

Connecticut State Department of Education

Abstract

This paper describes a new, portfolio-based assessment program for beginning science teachers who seek a license to teach in Connecticut's middle and high schools. The program is part of the comprehensive Beginning Educator Support and Training (BEST) induction program for beginning teachers in Connecticut. The portfolio-based assessment program was developed and pilot-studied during the 1993-1996 school years. The science portfolio assessment requires all beginning science teachers to plan, teach and assess students' learning in a two-week inquiry-based unit. The completed portfolios contain four types of artifacts: Daily logs for the two-week unit, written by the teacher; four videotaped segments that capture teaching and learning in various learning settings in the same unit; students' work done through the entire unit; and teacher's reflection on the quality of students' learning and his/her own teaching. Evidence gathered through various explorative studies of this new assessment system support the job-relatedness of the assessment content and the validity of inferences made from the assessment scores. Research results also show that the portfolio-based assessment program is accessible to beginning science teachers in different school systems and it has no adverse impact on any sub-group of candidates. Follow-up interviews with "graduates" of the program showed that the BEST program for science teachers, through its combination of licensure accountability testing and the induction's support component, has a lasting impact on many science teachers who are involved in the program, thus supporting the current state's systemic efforts to reform science education in its schools.

The Validity of Portfolio-Based Teacher Assessment

Michal Lomask, Michael Seroussi & Fie Budzinski

Introduction

The Beginning Educator Support and Training (BEST) program was designed by the Connecticut State Department of Education (CSDE) as an induction program for all beginning teachers who seek a Provisional Educator teaching certificate in Connecticut. The BEST program combines support and opportunities for teachers' professional growth with licensing-related assessment activities. The BEST program for beginning science teachers was developed by the CSDE together with a group of practicing science educators from public schools and institutions of higher education in Connecticut. The BEST program for beginning science teachers in Connecticut includes the following support and assessment components:

I. Support Activities:

School Mentorship: - All beginning science teachers are assigned a school-based mentor or a mentor team for at least one year, to provide them with support in their initial steps of teaching. Mentoring is provided by experienced teachers who are trained to support the work of the beginner teacher. The support is given during school time, and it includes collegial classroom observations and post-observation discussions between the mentor and the beginning teacher. The program is supported by the CSDE, through small professional development funds to schools with beginning teachers.

Teaching Clinics: - The teaching clinics are offered, free of charge, to all beginning teacher during their first year of teaching. There are four different teaching clinics, which focus on general pedagogy skills, such as lesson planning, classroom management, instructional flow and monitoring students' learning progress.

Lab Safety Activity - The lab safety activity includes participation in an interactive videodisk (IVD) simulation of a science laboratory (Lomask, Jacobson and Hafner, 1995) in which candidates have to monitor, in real time, student work in a simulated lab, and identify various safety problems. Teachers' performance in the IVD safety simulation is recorded by the system and transferred to a regular videotape, which later is evaluated by safety experts. A follow-up safety workshop offers teachers opportunities to address various issues of lab safety practices and local school regulations.

Regional Science Support Seminars - A series of six seminars is offered, free of charge, to all beginning science teachers, in four regional educational centers in the state. Through modeling, discussions and analyses of teachers' and students' work in science, beginning science teachers collaboratively explore issues of school science learning, such as teaching for understanding, encouraging enquiry and assessing students' academic performance. In addition, the science seminars focus on the development of reflective teaching practices and documentation of practice through teaching portfolios.

II. Licensing Assessment Activities

Assessment of general pedagogical skills - During the first year of teaching, beginning science teachers focus on the development of basic teaching competencies and good instructional habits. Their teaching competencies are evaluated through classroom observations by proficient state assessors, using the Connecticut Competency Instrument (Pecheone & Carey, 1989).

Assessment of content-specific pedagogy - During the second year of teaching, science teachers focus on the integration of content and instruction. Teachers are encouraged to identify main conceptual foci for their teaching, engage students in extended inquiries, evaluate students' learning, and reflect on the quality of their own performance. At the end of their second year of teaching, science teachers are required to submit a comprehensive portfolio of their work (Lomask, Baron & Pecheone, 1995). The quality of teachers' performance, as documented in their portfolios, is evaluated by highly-trained science teachers, who have to decide if the performance meets a pre established set of performance standards.

As can be seen from the above description, the BEST program combines the accountability aspect of licensing assessment with the support and professional development that beginning teachers need during the critical period of their induction into the teaching profession. The rest of this paper will describe the portfolio-based teacher assessment and explore validity-related aspects of this licensing examination instrument.

Developing Professional Science Teaching Standards

The development of licensing testing program should start with the conceptualization of the purpose of the assessment. In the BEST program the purpose of the licensing assessment was defined as ensuring that beginning teachers have mastered the knowledge, skills and behaviors necessary to teach their students effectively. To define what beginning teachers have to know and be able to do to teach effectively, there is a need to explore the factors that affect the work of beginning teachers. Reynolds (1992), in an analysis of the literature about competent beginning teaching concludes that beginning teachers oftenly do not know their subject matter in a way that allows them to explain it to students. Not having much experience with students, beginning teachers are unable to tailor materials and instruction to individual students. Unlike experienced teachers, who already have developed a rich and connected schemata of professional knowledge which allows them to respond to various classroom situations (Lienhardt, 1990), beginning teachers have less developed schemata, and they are challenged almost every day by the unfamiliar demands of classroom teaching. To perform their job beginning teachers have to not only develop the skills of classroom management and delivery of instruction, but also to develop a deep understanding of the discipline they teach and the various ways in which their students learn this discipline (Shulman, 1986). On the other hand, in most American schools beginning teachers are not working as interns, but rather they have full responsibility for the learning of their students, like all other teachers in the school (Interstate New Teacher Assessment and Support Consortium (*INTASC*), 1994). The implications of this reality are that beginning teachers who apply for a permanent teaching license (usually at the end of the second year of teaching), have to meet acceptable performance standards and show that they can teach their students successfully.

To measure how well beginning science teachers perform, there is a need to articulate what all science teachers are expected to know and be able to do, in the form of professional science teaching standards. The BEST's Professional Science Teaching Standards (PSTS) were developed by a committee of science educators from various Connecticut schools and institutions of higher education, to ensure a wide representation of subject-matter experts in the process. The committee was informed by current publications about science education, such as *Science for All Americans* (American Association for the Advancement of Science, 1989, 1994), an early draft of the *National Science Education Standards* (National Research Council, 1994), and a draft of the *Standards for Adolescence and Young Adulthoods/Science*

Certification (National Board for Professional Teaching Standards, 1993). The final version of the professional science teaching standards (see Appendix 1) contains four major standards which describe, in broad strokes, the specific knowledge, skills and behaviors necessary for effective science teaching.

The professional science teaching standards are central to the licensing process as they serve as the basis for the major components of the assessment system (i.e., assessment's format and content; assessment's evaluation and scoring procedures; performance standards and cut scores). Because of their centrality to the licensing process, these standards were validated through a national survey of science education experts, as is described later in this paper.

Assessment Format and Content

Generally, in designing the specific tasks of a licensing assessment, one has to establish legitimate relationships between the requirements of the assessment and its purpose. In the case of the BEST science teacher licensing program we built these relationships by ensuring that the assessment tasks are:

Credible - elicits teachers' knowledge and skills which are considered to be relevant and important for effective science teaching by professionals in the field.

Contextual - represents the daily work of science teachers with their students, in specific schools, classes and programs.

Coherent - explicitly connects teaching performances to expected students' learning outcomes in science.

Accessible - can be understood and executed by diverse groups of practicing science teachers, in diverse educational settings.

Multi-modal - includes multiple methods of evidence collection, to provide a comprehensive portrait of the individual teacher's performance.

To create assessment that can address all of the above characteristics, it was decided to base the assessment on *self-documented teaching portfolios*. Self-documented teaching portfolios are portfolios that are developed by individual teachers and contain only those artifacts that are directly related to the teachers' work with their students. Although other assessment methods, such as classroom observations or assessment centers (Burry & Oxford, 1995; Long & Stansbury, 1994) might address the same concerns, the portfolio-based teaching assessment offers several advantages:

• Teaching portfolios provide authentic presentations of the teacher's work over time. Through their portfolios teachers can present the rationale for their work, document significant teaching and learning events, collect and analyze students' work, and create a permanent record of their professional school performance. As such, the portfolio provides richer and more comprehensive data for the assessment process.

• Teaching portfolios are linked to professional teaching standards, without promoting standardized teaching practices. Since each portfolio developer describes the teaching that occurred within his/her specific context, the portfolios are evaluated in light of the unique teaching circumstances of their developer. In addition, teachers are asked to not only describe and analyze their teaching, but to also support their statements with evidence from students' work. As such, the portfolio-based assessment promotes a wide range of individual and contextualized responses to a single set of prescribed tasks and standards.

• Developing a teaching portfolio is a learning experience for the teacher. Intellectual and emotional involvement in the development process are likely to invite various reflective practices, such as discussions with peers, selective examination of materials and focused writing. Writing has the potential to transform vague ideas into clear concepts, thus promoting critical thinking and deepening teachers' understanding of their profession.

• Teaching portfolios can have multiple applications. Portfolios are lasting products that can be used by teachers in a variety of situations, beyond the licensure assessment. The portfolio, for example, can be shared with peers and serve as a basis for staff development and curriculum revisions. The portfolio, which is an authentic representation of the teacher's professional performance, can also benefit its developer when seeking new professional career opportunities.

Currently, all beginning science teachers in Connecticut have to develop and submit for evaluation a complete science teaching portfolio during the second year of their teaching. The general structure and content of the BEST program's science teaching portfolio and its task-specific documentation is described in Appendix 2. As can be seen, the portfolio entries include teacher's daily logs, students' work, videotapes of actual teaching and learning and reflective commentary by the teacher.

The content of the portfolio tasks has been carefully designed to engage beginning science teachers in teaching activities that promote students' development of science literacy. The tasks were also crafted to elicit appropriate and sufficient evidence to inform the portfolio evaluation process. The specific content of the portfolio tasks (e.g., open-ended lab investigation, discussion of science-technology-society issue) is highly related to the goals of students' learning in science, as articulated by the Connecticut Academic Performance Testing (CAPT) science program (Lomask ,Baron & Greig, 1997). Appendix 3 describes and compares the CAPT program of student assessment and the BEST program for the assessment of science teachers in Connecticut. As can be seen from this table, the two programs are aligned to be coherent and promote students' learning of science. More details about the science teaching portfolio can be found in the publication *Handbook for the Development of a Science Teaching Portfolio* (Lomask, 1996).

Evaluation and Scoring of Teachers' Performances

Portfolio-based assessment of science teachers, like other performance-based assessment programs, usually result in high credibility within the field, but has also experienced some psychometric challenges. Messick (1994) suggests that the two major threats to the validity of performance-base assessment are construct irrelevant variance (measuring something other than the intended knowledge and skills) and content under-representation (narrow sampling of the measured domain). To address these concerns a comprehensive, but highly structured evaluation and scoring framework, was developed.

The assessment process starts with a careful reading of the portfolio by a pair of trained assessors. The assessors, individually, review the portfolio materials and organize relevant evidence from the whole portfolio in tables. The tables are structured around criteria related to the PSTS. For example, the table that is used to review the daily logs organizes specific evidence related to the conceptual progression of the unit (Standard II.1), opportunities for students' explorations of ideas (Standard I.2 & II.2) and management of instructional resources (Standards III.2). Other tables address other standards. In this way only evidence for relevant performance are recorded and later used for the evaluation process. Based on evidence organized in the tables from the entire set of portfolio entries, the assessors answer a series of 12 guiding questions, designed to elicit specific patterns of teaching performances. The evidence and the identified patterns are then combined together, to describe how the teacher performs the following three major tasks:

I. Setting appropriate expectations for students' learning
II. Creating supportive environment for students' science learning
III. Evaluating students' learning and informing teaching

For each one of the above categories the assessors assign an intermediate score (on a 1-5 scale), based on evidence collected from all the tables, thus ensuring representations of all the measured standards. The three intermediate scores are combined by the two assessors who, together, must discuss and arrive at an agreed-upon decision about the quality of the overall teacher's performance. A 5-point rubric that incorporates the language of the professional science teaching standards serves as a *decision guide* for the final portfolio score and also as a template for a performance feedback for the candidates (the general flow of the evaluation and scoring process is described in Figure 1.) Benchmark portfolios were identified and they are used to train new portfolio assessors in the structured process of portfolio evaluation.

It is important to note that the portfolio scoring criteria are described in the *Handbook for the Development of a Science Teaching Portfolio* , so that beginning science teachers and their mentor have the opportunity to share and discuss the evaluation framework well in advance. In addition, before the submission of the portfolio, beginning teachers are encouraged to evaluate their own performance, in the same way it will be evaluated later by state assessors. This provides beginning teachers with opportunities to self-evaluate the quality of their teaching before they submit their portfolios for the formal state assessment.

As of today, three "generations" of beginning science teachers have participated in the pilot studies of the BEST assessment program. A total of 223 science portfolios were developed, submitted and scored, and about fifty experienced science teachers were trained as portfolio assessors or as leaders of support seminars for beginning science teachers. The program was introduced to staff at the higher education institutions that prepare science teachers in Connecticut and several of the programs are already implementing some of the portfolio tasks with their graduates.

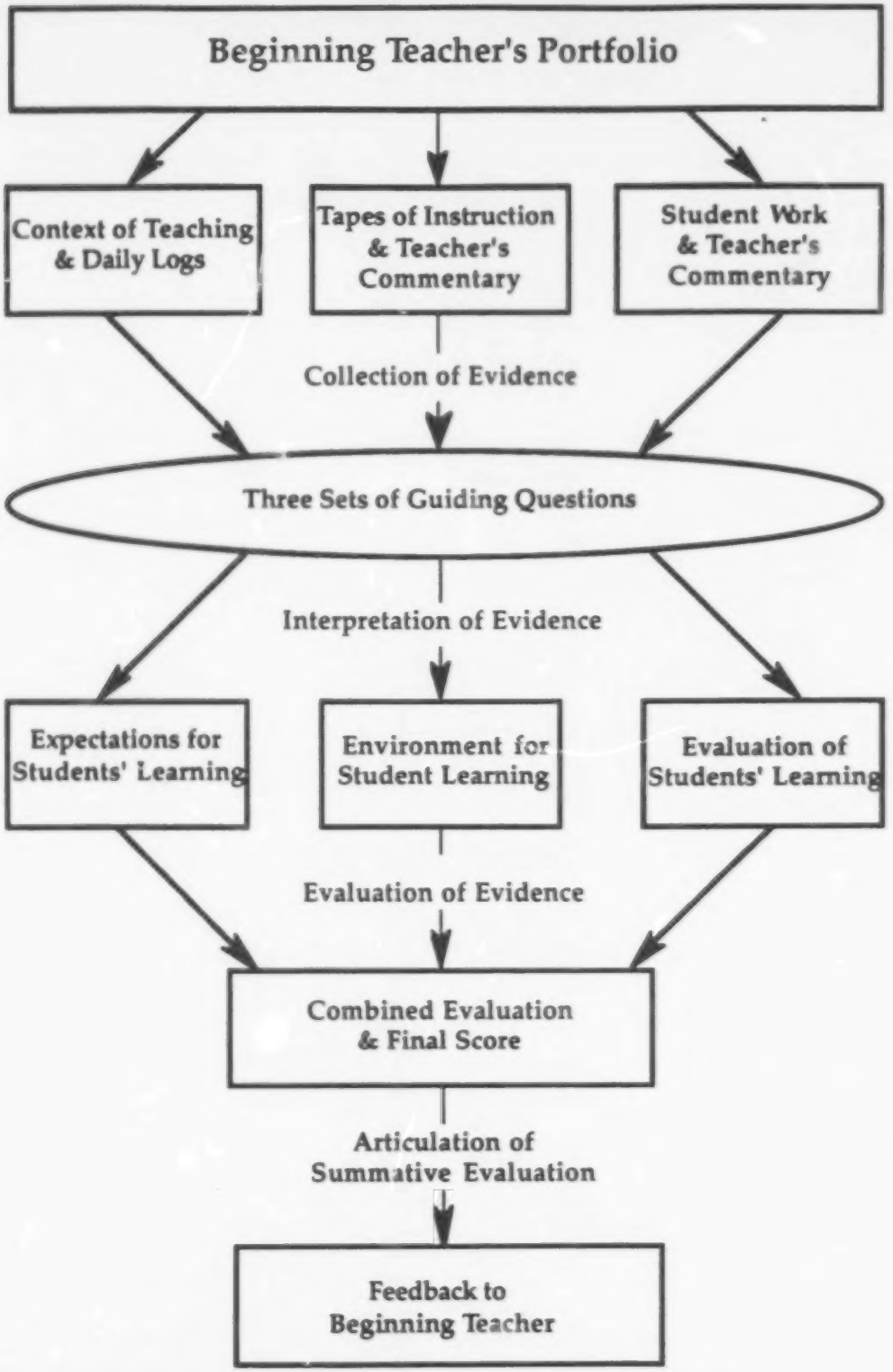
Beginning Teacher's Portfolio
Context of Teaching & Daily Logs
Tapes of Instruction & Teacher's Commentary
Student Work & Teacher's Commentary
Collection of Evidence
Three Sets of Guiding Questions
Interpretation of Evidence
Expectations for Students' Learning
Environment for Student Learning
Evaluation of Students' Learning
Evaluation of Evidence
Combined Evaluation & Final Score
Articulation of Summative Evaluation
Feedback to Beginning Teacher

Figure 1: A model for the evaluation of science teaching portfolios

Studies Exploring the Validity of the Portfolio-based Science Teacher Assessment

As was described before, the portfolio-based assessment of beginning science teachers is part of the licensure process of beginning teachers in Connecticut. The *Standards for Educational and Psychological Testing* (AERA et al, 1985) state that the purpose of the licensure testing is to protect the public. In the case of licensing teachers, "protection of the public" means protecting students and ensuring that they are given fair opportunities to attain specific educational goals, as defined by their school curriculum. The licensure assessment of beginning teachers should be designed to distinguish between those teachers who do and those who do not posses the knowledge, skills and behaviors believed to be necessary to perform their job well at the time of entry into the teaching profession. How should performance-based licensure assessment be validated?

As in other licensure tests, the validity of the teacher licensure program rests on the appropriateness, meaningfulness, and usefulness of inferences made from test scores (AERA et al, 1985). These aspects of the program should be judged through a continuous study of the meaning of the test requirements, the use of test scores (Messick, 1994, 1995), and the consequences associated with test score interpretations (Haertel, 1987). Kocher and Tannenbaum (1996), based on review of the literature about licensure testing, recommend that the validation of licensure testing be based on content-related studies, with the involvement of subject-matter-experts in all stages of the test development. The BEST program for science teachers indeed was developed by expert science educators who participated in all stages of the program, from the initial development of professional science teaching standards, through the development of portfolio tasks, to the final setting of scoring criteria and performance standards. The following sections describe some of the studies conducted to explore validity-related issues of the BEST portfolio-based licensure assessment.

Evaluation of the Professional Science Teaching Standards (PSTS)

The PSTS were developed by a group of Connecticut science educators. To evaluate the quality of these standards and their credibility in the field, the PSTS document was mailed to one hundred science educators teaching in American universities (randomly selected from the *NARST Directory of Leadership and Members*, 1994) and one hundred local science teachers, randomly selected from the state data base. The reviewers were asked to use a five-point Likert scale to express their agreement with statements about the qualities of the standards. Eighty five questionnaires

(42%) were returned and analyzed. The results (see Table 1) showed that all the standards were regarded important and appropriate for science teachers, and aligned with the vision of effective science teaching, as described in the current National Science Education Standards document (NRC, 1996).

Table 1

Means and standard deviations of rating (N=85) to survey questions (scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree)

	Mean & (SD)		
Standard	Q1* Importance	Q2* Appropriateness	Q3* Alignment
I.1	4.85 (0.39)	4.42 (0.80)	4.49 (0.68)
I.2	4.81 (0.42)	4.42 (0.77)	4.05 (0.94)
I.3**			
II.1	4.84 (0.37)	4.50 (0.76)	4.46 (0.63)
II.2	4.54 (0.61)	4.21 (0.86)	4.05 (0.79)
II.3	4.75 (0.55)	4.48 (0.82)	4.30 (0.74)
III.1	4.80 (0.46)	4.40 (0.82)	4.35 (0.81)
III.2	4.81 (0.34)	4.46 (0.77)	4.38 (0.79)
III.3	4.74 (0.47)	4.33 (0.89)	4.36 (0.69)
III.4	4.79 (0.56)	4.47 (0.73)	4.40 (0.75)
IV.1	4.71 (0.61)	4.33 (0.88)	4.40 (0.78)
IV.2	4.68 (0.63)	4.30 (0.95)	4.40 (0.81)
IV.3	4.21 (0.78)	3.79 (0.96)	3.77 (0.91)
Total	**4.68 (0.39)**	**4.25 (0.78)**	**4.26 (0.51)**

* Q1: The standard describes an important aspect of science teaching
Q2: The standard represents appropriate expectations for beginning science teachers
Q3: The standard is aligned with the vision of the National Science Teaching Standards

** S1.3 was added after this survey, based on reviewers" recommendations

Evaluation of the coherence of the assessment components

Part of the construct validity of performance-based assessments is established through analyses of the connections among assessment purpose (mastery of knowledge, skills and behaviors as described in the PSTS, in this case) and the assessment requirements (the specific portfolio tasks and evaluation framework). To evaluate the coherence of the assessment system, fifty experienced science educators were asked to review the complete set of assessment materials (i.e., PSTS document, the *Handbook for the development of a science portfolio*, the manual for portfolio evaluation), and examine the technical quality (i.e., clarity, organization), and the internal relations among the different components. The reviewers were asked to use a five-point Likert scale to express their agreement with various evaluative statements. Thirty-four teachers (68%) completed the review and answered all questions. Their responses are summarized in Table 2.

Table 2

Means and Standard Deviations of Rating (N=34) Review Questions (scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree)

Review Statements	Means and SD
The teaching standards are written in a clear and concise language	4.47 (0.62)
The teaching standards focus on important aspects of science teaching	4.82 (0.59)
The portfolio tasks are written in a clear and concise language	4.76 (0.56)
The portfolio tasks are connected to students' learning goals	4.71 (0.59)
The portfolio tasks represent important teaching activities	4.82 (0.42)
The portfolio assessment tasks reflect the teaching standards	4.72 (0.47)
The portfolio scoring framework reflects the teaching standards	4.86 (0.39)

As can be learned from results presented in Table 2, the assessment documents were perceived as organized and well written. More importantly, experienced science teachers, without direct guidance, were able to see the connections between the professional science teaching standards, the portfolio-based assessment tasks and the evaluation framework for the science portfolios, which supports the coherence of the assessment system.

Evaluation of the fairness of portfolio assessment

The concept of fairness in testing has been most closely associated with the process of assuring that the content, substance, and language used in examination and examination-related materials are not offensive or inappropriate for the diversity of minority and /or ethnic groups represented in the candidate population (Educational Testing Service, 1980). The Standards for Educational and Psychological Testing (1985) states that test developers should consider the content and type of assessment in relation to cultural backgrounds and prior experiences of the variety of ethnic, cultural, age, and gender groups represented in the intended population of test takers. Bond and Linn (1996) who studied assessment practices of the National Board for Professional Teaching Standards (NBPTS) extended the definition of fairness to include both content analysis and relevant empirical data. To address both aspects of fairness review, two studies have been conducted.

In the first study, conducted by the Professional Examination Service (PES) department of research and development (Smith and Greenberg, 1996), the review focused on assessment matters related to the content, language, and access to resources. A committee of fifteen science subject matter experts (five female, ten male; seven minorities, nine whites; two higher education, three science department chairs, three administrators, two beginning teachers, two middle school teachers, three high school teachers) was convened by the PES. After formal presentations by PES and CSDE staff, members of the committee reviewed the materials and unclear issues (e.g., support provided to beginning teachers) were clarified. Finally, committee members were asked to complete a set of ten fairness review questions (see Appendix 4) with a yes/no response. Analysis of the rating indicated that the reviewers strongly supported the fairness of the science portfolio assessment (93% of members answered with *yes* to nine out of the ten items. Only item 6, which deals with issues not related to science teaching, received 60% of *yes* responses).

In the second study, conducted by the authors, the review focused on comparisons of empirical data about actual teachers' performances. For this study the scores given to 57 beginning science teachers, who participated in a pilot study of the science portfolio assessment, were analyzed by background factors (i.e., gender of portfolio developer, school level, subject matter of portfolio, and socioeconomic status of school community, as determined by the CT Economic Reference Group). The analyses did not reveal any significant effect of any of these factors on teachers' portfolio scores (see figures 1 to 4).

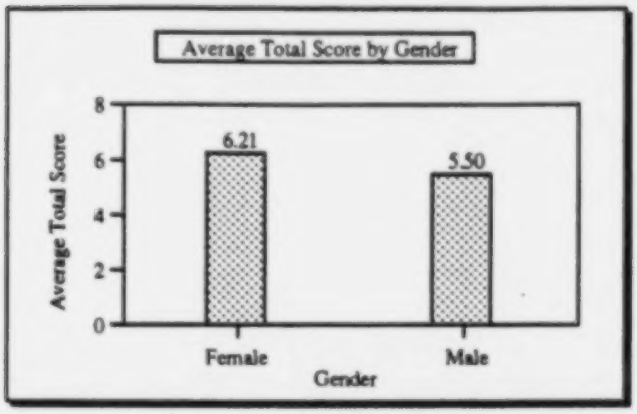

Figure 1. *Average portfolio score, by gender*

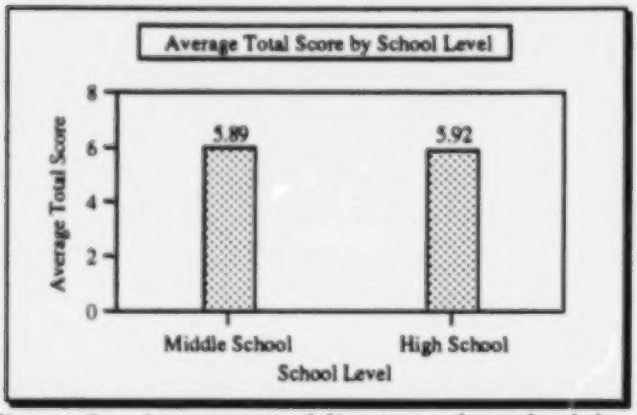

Figure 2. *Average portfolio score, by school level*

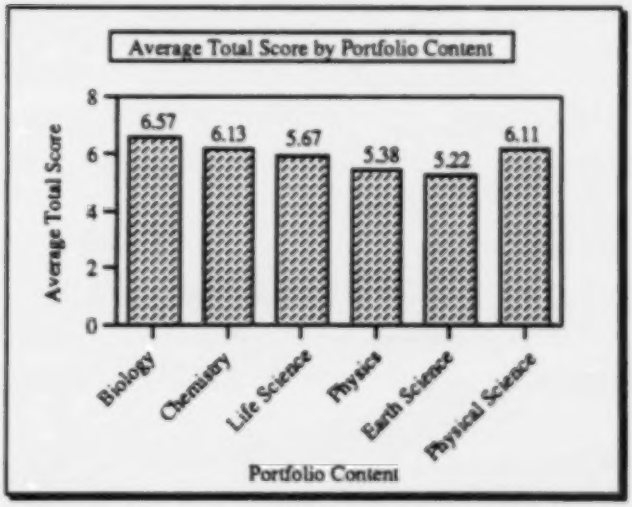

Figure 3. *Average portfolio score, by subject matter*

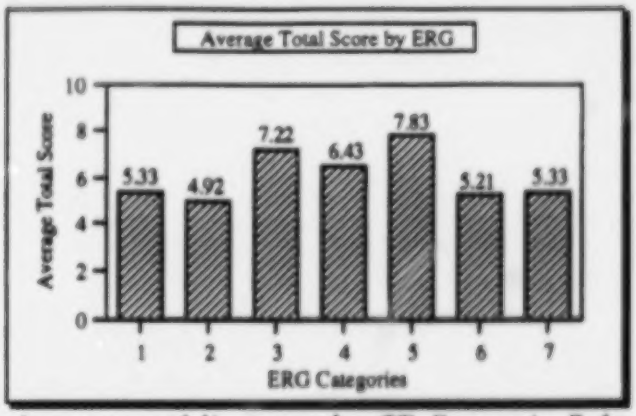

Figure 4. *Average portfolio score, by CT Economic Reference Group*

The only significant difference between the mean scores of candidates sub-groups was between teachers in ERG 5 and teachers in ERG 2. But since the total number of teachers in each economic group was small (three to nine teachers in each group), these findings bear no significance. It is important to mention here that the effect of candidates' race on performance was not explored in this study, since Connecticut did not have any minority beginning science teachers in the years in which the study was conducted (the possible reasons for this situation are beyond the scope of this paper). The findings of this study, which indicate that the portfolio-based assessment has no adverse impact on any sub-groups of candidates, further support the fairness-related validity aspect of this assessment.

Internal consistency and reliability of the scoring process

There are different ways to estimate the validity and reliability of assessment scores. In the case of licensing examinations, the estimation of reliability focuses on the reliability of licensing decisions that are made by the assessment. The portfolio-based assessment, which consumes a considerable amount of candidates' and assessors' time, challenges traditional approaches to the study of reliability, such as test-retest, or alternate form reliability. To study the internal consistency and the reliability of the portfolio assessment, two types of score analyses were performed. In the first analysis, the internal consistency of the assessment was explored by calculating the correlations among the intermediate scores that were given to candidates on each of the three evaluation categories (see page 9). Table 3 summarizes the findings which show that scores for the three categories are moderate to highly correlated, and contributing equally to the final portfolio score. These results were expected, as the three tasks (planning, teaching and assessment) are different but are based on similar components of knowledge and skills.

Table 3

Pearson product correlations among scores for the science portfolio evaluative categories

	Total portfolio score	Planning Learning	Facilitating Learning
Planning Learning	0.86*		
Facilitating Learning	0.84*	0.73*	
Evaluating Learning	0.83*	0.65*	0.66

* $p<0.01$

The second analysis focused on the consistency of scores given by different pairs of assessors. For this study thirty portfolios were scored twice, each scored by two independent pairs of trained assessors who assigned a final score of 1 to 5 to each portfolio. Since the science portfolios encompass five different certification areas (i.e., biology, chemistry, physics, Earth science and general science), and the assessors were certified and matched to the same subject matter of the evaluated portfolio, ten different pairs of assessors were needed in this study (each pair scored six portfolios.) Across all matched pairs, there was exact agreement on the final score for 19 out of the 30 scored portfolios (63%) and adjacent agreement on the scores for 28 portfolios (93%). These results were not considered sufficient for establishing consistent licensure decisions. However, it should be noted that these results were obtained with a group of science teachers who were trained to be portfolio assessors, but whose proficiency as scorers were not established. Once the assessment is on line, and cut off scores established, assessors will not have to identify accurately five levels of performance, as they did in this study, but rather judge if each portfolio's performance meets, or does not meet, established performance standard, which might minimize classification errors and increase reliability. Given the early stages of portfolio scoring development, this study showed that the established scoring system is promising, although more work is needed in order to improve the training of proficient assessors.

Discussion

This paper describes a new, portfolio-based assessment for beginning science teachers and several studies that were conducted to explore the reliability and validity of the assessment program. The use of teaching portfolios in education is not completely new. The Teacher Assessment Project (Collins, 1993) used teaching portfolios for the assessment of experienced science teachers and found the process to have high degree of fidelity to teaching and an excellent source of professional development for the participating teachers. Teaching portfolios are part of the assessment system for teachers who seek National Board Certification (NBPTS, 1997). Teaching portfolios for beginning mathematics and English language arts teachers are currently under development and being studied by various INTASC researchers (Moss et al, 1995). Their studies, like ours, focus on the continuous exploration of the meaning, interpretations and consequences of this assessment.

Connecticut, however, is the first state to use teaching portfolios as part of mandated licensure requirements. This aspect challenged us to show that in addition to its fidelity to the field, and its potential impact on teachers and students, we can also establish its legal defensibility. Data collected through several studies demonstrated that the science portfolio-based assessment is credible, contextualized, coherent, fair and accessible for beginning teachers, thus supporting the content validity of this assessment. Additional support for the validity of this assessment comes from its congruence with the state-wide assessment of high school students' learning of science (Appendix 3.) The two programs, the student assessment and the teacher assessment, support each other and help to build a coherent, standards-base system of science education, in which all students have the opportunity to develop science literacy. The established expert-based scoring system supports the validity of the assessment and promises to produce reliable scores for future licensure decisions

What do beginning teachers think about this new assessment? Written feedback from beginning teachers who participated in the pilot studies indicated that the process of portfolio development is very time consuming for most teachers. In average, teachers spent forty to sixty hours on data collection, documentation and reflective writing. But the feedback also indicated that many beginning teachers considered the process of portfolio development, and the program's support seminars, as an opportunity for reflection and professional growth. This sentiment is sincerely expressed by the words of a beginning teacher who volunteered to participate in one of the pilot studies:

> I selected to participate in the pilot study because it replaced the traditional assessment by classroom observations. Half way through the process, while I was writing down daily logs and collecting students' work, I was kicking myself: Why did I choose to do it? After I've completed the portfolio, I was happy that I chose it. The portfolio is such an accurate reflection of what you're doing in your classroom. You can't lie on videotape. You can see if there was any learning in the classroom, if there was any progress.

There are still several aspects of the program that need to be explored. For example, the issue of equity is a major challenge for any performance-based assessment (Linn & Baker, 1996), including this assessment. Differences in the extent to which schools support the current trends of reform in science education might affect the level of support that beginning teachers receive in their efforts to develop and implement new curriculum and instructional methods. To address this issue, the BEST program offers support seminars to all beginning science teachers, their school mentors and regional lead science teachers. During these sessions teachers are engaged in various explorations of the science teaching standards and are encouraged to find ways to implement them in their own school context. The support is given during the year and a half that precede the portfolio development, thus providing teachers with access to information and expertise needed to perform the portfolio's tasks.

Another issue of concern, which is now under study, is the effect of teachers' writing ability on their perceived teaching ability. For example, some portfolios, through the videotaped segments, show science teachers who create an interesting and challenging learning environment in their classes, engaging students in thought-provoking problems. The writing of some of these same teachers, however, is limited and presents a non-reflective type of teaching practice, which might affect the final score for the whole portfolio. The plans, the description of activities, the videotapes and student work can all compensate for lack of articulation, but still this issue deserve a thorough study.

To summarize, the portfolio-based licensure assessment is based on authentic teaching activities and it models needed school curriculum reform. Through continuous support and staff development, the program reaches all beginning science teachers and their school mentors, thus promising to be not only a valid tool for licensing decisions, but also a lever for the improvement of science instruction in the state public schools.

References

American Association for the Advancement of Science (Project 2061), (1989). Science for all Americans. Washington, DC: American Association for the Advancement of Science publication.

American Association for the Advancement of Science (Project 2061), (1993). Benchmarks for Science Literacy. Washington, DC: American Association for the Advancement of Science publication.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1985). Standards for Educational and Psychological Testing. Washington DC: American Psychological Association.

Bond, L., & Linn, R. (1996). Documenting the extensiveness and etiology of adverse impact. A paper presented at the annual meeting of the American Educational Research Association, New York, New York.

Burry, J. A. & Oxford, R. (1995). Expert Science Teaching Educational Evaluation Model (ESTEEM) for measuring excellence in science teaching for professional development. A paper presented at the annual meeting of the National Association of Research in Science Teaching, San Francisco.

Collins, A., (1993). Performance-based assessment of biology teachers: promises and pitfalls. Journal of Research in Science Teaching, 30, 1103-1120.

Educational Testing Service (1980). An approach for identifying and minimizing bias in standardized tests: A set of guidelines. Princeton, NJ: Author.

Haertel, E. (1987). "Validity of teacher licensure and teacher education admission tests". Paper prepared for the National Education Association and Council of Chief State School Officers.

Interstate New Teacher Assessment and Support Consortium (1994). Model Standards for Beginning Teacher Licensing and Development: A Resource for State Dialogue. Council of Chief State School Officers, Washington, DC.

Leinhardt, G. (1990). Capturing the craft knowledge in teaching. Journal of Educational Psychology, 78, 75-79.

Linn, L. R., & Baker, L. E. (1996). Can performance-based student assessment be psychometrically sound? In J. B. Baron & D. P. Wolf (eds.), Performance-Based Student Assessment: Challenges and Possibilities. Ninety-fifth Yearbook of the National Society for the Study of Education, University of Chicago Press.

Lomask, S. M., (1996). Handbook for the Development of a Science Teaching Portfolio. Published by the Connecticut State Department of Education.

Lomask, S. M., Jacobson, L., & Hafner L. (1995). The development and validation of an assessment of beginning science teachers' knowledge of lab safety, using interactive videodisk technologies. Science Education, 79, pp. 519-534.

Lomask, S. M., Pecheone, L. R., & Baron B. J. (1995). Assessing new science teachers. Educational Leadership, 52, 62-65.

Lomask, S. M., Baron, B. J., & Greig , J. (1997, in press). Large scale performance assessment in Connecticut: Challenges and Resolutions. In B. Fraser and K. Tobin (eds.), International Handbook of Science Education, Kluwer, The Netherlands.

Messick, s. (1994). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23, 13-23.

Messick, S. (1995). Validation of inferences from persons' responses and performances as scientific inquiry into meaning. American Psychologist, 50, 741-749.

Moss, P. A., Pecheone, L. R., Pence, L. & Thompson, W. (1995). A preliminary validity research agenda for the INTASC Beginning Teacher Assessment. Unpublished manuscript, Interstate New Teacher Support and Assessment Consortium and the Council of Chief State School Officers, Washington, DC.

National Board for Professional Teaching Standards (1993). Adolescent and Young Adulthood/Science Standards for National Board Certification (a draft). NBPTS: Author.

National Board for Professional Teaching Standards (1997). The Backgrounders. NBPTS: Author.

National Research Council (1994). National Science Education Standards (a draft). Washington DC: Author.

Pecheone, R. L., & Carey, N. B. (1989). The validity of performance assessment for teacher licensure: Connecticut's ongoing research. Journal of Personal Evaluation, 3, 115-141.

Reynolds, A. (1992). What is competent beginning teaching? A review of the literature. Review of Educational Research, 62, 1-35.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15, 4-14.

Smith, I. L., & Greenberg, S. (1996). Fairness review of CSDE's mathematics and science portfolios. Final Report, Professional Examination Service: Author.

Stansbury, K. &

United States Department of Health, Education and Welfare (1971). Report on licensure and related health personnel Credentialing. PHEW Publication 72-11, Washington, DC: Author.

Appendix 1

The Professional Science Teaching Standards

Standard I: Science teachers understand and respect student diversity and provide all students with opportunities to learn important and challenging science.

I.1 Students' Diversity
Science teachers believe that all students have the potential to learn and be successful in learning science. They search for science materials and teaching strategies that encourage students with diverse abilities, interests and backgrounds to actively participate in the learning of science.

I.2 Students' Learning
Science teachers understand how students' prior knowledge of science affects how they construct new knowledge. They explore students' thinking, respect their ways of knowing and strive to facilitate students' confidence in their ability to learn new concepts, solve problems and make informed decisions.

I.3 Students' Culture
Science teachers understand the cognitive, social and emotional development of students and how it affects their learning. They familiarize themselves with students' cultures and encourage students to assume an active and responsible role in their own learning of science.

Standard II: Science teachers have a thorough understanding of the content and nature of science and the relationship of science and technology to society.

II.1 Science Nature and Content
Science teachers understand the nature and content of the various science disciplines. They focus students' learning on the major concepts, theories and questions of science by engaging them in thoughtful explorations of the physical world.

II.2 Science Logic and Construction of Knowledge
Science teachers understand the contributions of various forms of scientific thinking to the practice of science. They create opportunities for students to develop thinking patterns that are essential for the development of students' independent, creative and critical reasoning.

II.3 Science Context and Applications
Science teachers understand the significance of scientific literacy for active participation in a modern society. They create opportunities for students to apply their knowledge to solve problems, examine science-related issues and construct informed and carefully reasoned opinions.

Standard III: Science teachers build learning environments that promote rigorous learning and encourage students to use inquiry and scientific habits of mind.

III.1 Learning Environment
Science teachers understand how to turn the science classroom into a science learning community. They create a physically and emotionally safe learning environment in which teacher and students are free to ask questions, seek information and validate explanations in various thoughtful, creative and cooperative ways.

III.2 Instructional Resources
Science teachers understand how to manage time, space and resources to support students' learning. They ensure the availability of materials, equipment and communication technologies to support students' scientific investigations.

III.3 Student Assessment
Science teachers understand how student assessment can support instruction and strengthens learning. They engage students in thoughtful assessments and provide them with informative feedback to enhance their ability to become independent learners.

III.4 Reflective Practice
Science teachers understand how to use assessment data to improve teaching and learning. They continuously gather data and explore patterns of performance in students' work, in order to improve students' learning and their own teaching.

Standard IV: Science teachers function as members of a larger, professional learning community.

IV.1 Continual Learning
Science teachers understand that being a teacher means being a scholar of human learning. They identify their own learning needs and take active steps to strengthen their knowledge of science and their understanding of students as learners of science.

IV.2 Collegiality
Science teachers understand that they are part of a community of learners. They seek the advice of other professionals and they contribute to the quality of their colleagues' practice and to the work of the larger educational community.

IV.3 Family and Community Outreach
Science teachers understand the significant influence of the larger community on student learning. They routinely involve families and other members of the community to best serve the interests of each student.

Appendix 2:

Summary of the Portfolio Tasks for Beginning Science Teachers

Task	What to Do	What to Submit
Describe a Unit of Inquiry-Based Science Learning	⇨ Select, adapt or develop a 2-week inquiry-based learning unit ⇨ Provide relevant information about students in class ⇨ Describe expected students' learning during the unit ⇨ Keep a daily log of classroom events and evaluation of students' learning	• Class & Community Profile form • An introduction, context & goals to the portfolio unit • Daily logs
Describe Students' Science Explorations	⇨ Videotape a lesson in which students are engaged in a lab- based inquiry, data collection & interpretation ⇨ Videotape a lesson in which students learn to analyze and discuss issues of science, technology and society ⇨ Analyze the quality of students' learning in these two instructional settings	• Videotaped segments of the lab activity • Videotaped segment of the STS activity • Commentary about student learning the lab activity • Commentary about student learning in the STS activity
Evaluate Students' Learning of Science	⇨ Collect the work that was done by two students during the entire unit ⇨ Evaluate the learning progress of these students ⇨ Analyze the effectiveness of your teaching, based on the learning of students in class ⇨ Suggest changes that you would make in your future teaching	• Collection of work done by two students • Detailed evaluation of these two students' learning • Commentary about your teaching and the resulting learning of students in your class
Evaluate the Quality of Your Teaching	⇨ Use the Guidelines for Portfolio Evaluation to review, evaluate and score your teaching performance, as it is documented in your portfolio	• Self-evaluation of portfolio • Holistic scoring of your portfolio

BEST COPY AVAILABLE

Appendix 3:

The CAPT science assessment and the BEST's science portfolio tasks

CAPT-Science Components	Science Teaching Portfolio Tasks
• **A written science test** Assesses students' understanding of science concepts and their applications	• **Design of a learning unit** Unit is focused on students' explorations of essential science concepts and their applications. • **Analyses of student work** Evaluation is designed to assess and strengthen students' understanding of major science concepts and processes.
• **Lab investigation** Assesses students' ability to perform scientific inquiry	• **Open-ended lab activity task** Task designed to encourage and strengthen students' development of science inquiry skills
• **Interdisciplinary task** Assesses students' ability to make informed decision and write clearly about Science-Technology- Society issues	• **Science-Technology -Society (STS) task** Designed to engage students in serious examination of STS issues that require application of scientific knowledge and informed decision making

Appendix 4:

Questions for Fairness Review of Portfolio Assessment Materials

(from Smith & Greenberg, 1996)

1. Are the tasks described in the science portfolio equally familiar to all beginning teachers (e.g., consider diversity issues related to gender, race, ethnicity, disability)?
   yes❑ no❑

2. Are the evaluation criteria for the science portfolio tasks equally familiar to all beginning teachers (e.g., consider diversity issues related to gender, race, ethnicity, disability)?
   yes❑ no❑

3. Are the tasks described in the science portfolio free of language and/or content that might offend any beginning teachers (e.g., consider diversity issues related to gender, race, ethnicity, disability)?
   yes❑ no❑

4. Are the tasks described in the science portfolio free of language and/or content which reinforce stereotype of beginning teachers (e.g., consider diversity issues related to gender, race, ethnicity, disability)?
   yes❑ no❑

5. Are the tasks described in the science portfolio and the evaluation criteria equally appropriate for all beginning teachers within the certification code (i.e., all levels, all grades of biology, chemistry, physics, earth science , general science)?
   yes❑ no❑

6. Are all significant components in the teaching/learning process measured by the tasks in the science portfolio?
   yes❑ no❑

7. Do the tasks and evaluation criteria in the science portfolio accommodate different teaching contexi (e.g., urban, rural and suburban)?
   yes❑ no❑

8. Do the tasks and evaluation criteria in the science portfolio accommodate different curricula?
   yes❑ no❑

9. Is there access to support structures for beginning teachers to develop the teaching competencies evaluated in the science portfolio?
   yes❑ no❑

10. Are there provisions to ensure that the resources, materials, and equipment required to complete the task of portfolio assembly are available to beginning teachers?
    yes❑ no❑